Exploring Legal Patent Citations for Patent Valuation

Shuting Wang
Department of Computer Science and Engineering
Pennsylvania State University
sxw327@cse.psu.edu

Zhen Lei
Department of Energy and Mineral Engineering
Pennsylvania State University
zlei@psu.edu

Wang-Chien Lee
Department of Computer Science and Engineering
Pennsylvania State University
wlee@cse.psu.edu

ABSTRACT

Effective patent valuation is important for patent holders. Forward patent citations, widely used in assessing patent value, have been considered as reflecting knowledge flows, just like paper citations. However, patent citations also carry legal implication, which is important for patent valuation. We argue that patent citations can either be technological citations that indicate knowledge transfer or be legal citations that delimit the legal scope of citing patents. In this paper, we first develop citation-network based methods to infer patent quality measures at either the legal or technological dimension. Then we propose a probabilistic mixture approach to incorporate both the legal and technological dimensions in patent citations, and an iterative learning process that integrates a temporal decay function on legal citations, a probabilistic citation network based algorithm and a prediction model for patent valuation. We learn all the parameters together and use them for patent valuation. We demonstrate the effectiveness of our approach by using patent maintenance status as an indicator of patent value and discuss the insights we learned from this study.

1. INTRODUCTION

Patent valuation, i.e., assessing the value of patents, is an important but challenging task for firm technology and innovation management. Patent citations have been widely used in patent valuation [19, 8, 6, 7] on the ground that patent citations provide "paper trails" of knowledge flows among patents. The fact that a patent cites a large number of prior patents (hereafter, *backward citations*) suggests that the patented invention has built upon "the shoulders of giants", i.e., a significant amount of prior knowledge. This implies that the invention has great **technological richness** (defined as the amount of prior knowledge a patent builds upon) and likely high technological quality and economic value. Similarly, when a patent is cited by a large number of subsequent patents (hereafter, *forward citations*), this indicates that the patented invention has led to a number of successful lines of innovation. Thus, the invention is likely to be of high **technological influence** (defined as the technological impact that a patent has on subsequent inventions) and thus highly economically valuable.

From such technological aspects, one might think that patent citations are similar to paper citations. However, patent citations are actually quite different from paper citations in significant ways. In particular, patent citations could be interpreted in two dimensions: (i) a technological one that is related to knowledge flows, and (ii) a legal one that is related to delimitation of patent scope. Let's consider the scenario where a patent application is under examination. The patent examiner needs to search for relevant prior art (prior inventions) to determine whether the invention is patentable based on its novelty and inventiveness in comparison to these prior inventions. Meanwhile, the examiner needs to determine the appropriate scope of the patent right by asking the applicant to modify, if necessary, the language of the claims. For example, if an inventor applies for a chemical compound that makes some novel structural modification to an existing drug. The examiner would grant a patent to the invention, but cite the prior patent on the existing drug to: (i) show the knowledge link between the two inventions and (ii) to narrow down the scope of the newly granted patent so that it would cover only the modification, not the original chemical structure. In fact, the scope of the newly granted patent could be so narrowed down by the prior patent on the existing drug that the firm owning the newly issued patent may have to get a license from the patentee of the existing drug patent in order to market the new drug. In this case, the cited prior patent acts as a blocking patent to the newly granted patent. Similarly, an applicant, under the U.S. Patent Law, has the obligation to disclose relevant prior art that she knows during her research, though she has no obligation to identify all possible relevant prior art when filing an application. These applicant-inserted citations could, on one hand, suggest knowledge flows, but on the other hand, be used to narrow down the patent scope.

Therefore, a citation made by patent A to patent B could suggest that there are knowledge flows from the cited patent to the citing patent (this aspect of patent citations defines the *technological citations*), or that the cited patent puts legal constraints on the scope of the citing patent (this aspect defines the *legal citations*), or both. The legal interpretation of patent citations has quite different implication in terms of what patent citations mean for patent valuation, compared to the technological interpretation of patent citations. From the legal aspect, when a patent cites a large number of backward citations, it could suggest that many prior patents

CIKM'14, November 3–7, 2014, Shanghai, China.

http://dx.doi.org/10.1145/2661829.2662029 .

might have been used to narrow down the scope of the citing patent or block the citing patent. Consequently the citing patent might have a very narrow **legal patent scope** (i.e., the scope of patent right claimed), and thus likely small commercial value. On the other hand, when a patent receives a large number of forward citations, it might be blocking or putting constraints on these subsequent patents. In this case, a patent with a large number of forward citations implies a high level of **legal blocking power**, and thus it is a highly valuable patent.

Furthermore, legal constraints in patent citations (from the legal aspect) are time sensitive. For a patent citation with two year lag between the citing and cited patents, the cited patent could block the citing patent for a long time. However, if a patent cites an expired prior patent, then the cited patent put no legal constraints on the citing patent.

Our study aims to explore the insights about the technological and legal dimensions of patent citations and propose corresponding measures for patent valuation. Specifically, based on the technological and legal interpretations of patent citations, we propose to capture four quality measures of patents, namely, *technological richness*, *technological influence*, *legal patent scope*, and *legal blocking power*.

More importantly, and quite intuitively, we propose that there exist mutual interdependence among these four measures of patents; and the two measures in the technological dimension are related to each other in a different way than the other two measures in the legal dimension are related. Consider the technological richness and influence associated with a patent A. If patent A cites prior patents that are of greater technological influence, other things being equal, the technological richness of the citing patent A is greater, as the invention builds on a lot of influential prior inventions. Meanwhile, if patent A is cited by subsequent patents that are of high technological richness, the patent A's technological influence would be greater, as it leads to subsequent innovations of high quality.

By an interesting contrast, if patent A cites prior patents that are of greater legal blocking power, other things being equal, the legal patent scope of the citing patent A is smaller, as it is narrowed down by prior patents with large blocking power. However, if patent A is cited by subsequent patents that are of large legal patent scope, the patent A's legal blocking power is greater as it put constraints on subsequent patents with broad patent right that are highly valuable.

We investigate different methods to quantify the four proposed measures. We first assume that a patent citation can be interpreted in the technological dimension or in the legal one (or both). We then consider the case where a patent citation represents a probabilistic mixture of both technological and legal citations, with the significance of legal citations decaying by time (i.e.,the grant lag between a cited and citing patents). Accordingly, we capture their mutual interactions and iteratively learn the four measures using the patent data. Technically, we adopt a parameter learning process that integrates multiple models (including a temporal decay in legal citations, a probabilistic citation network based algorithm for quantifying the four proposed patent quality measures, and a prediction model for patent valuation).

To validate our idea of distinguishing legal citations from technological citations, we empirically apply the four proposed patent quality measures in patent valuation. We use patent renewal status (patent maintenance) as an indicator of patent value in our experimentation. Our results show that separating technological and legal dimension in patent citations achieves better accuracy in experiments for patent value prediction. And our proposed patent quality measures based on legal citations show more important roles in predicting patent value than measures based on technological citations. Our study also confirms the mutual interdependence between technological influence and technological richness is different from that between blocking power and legal patent scope. Moreover, by applying a probabilistic model to quantify the proposed concepts, we validate that patent citation is a probabilistic mixture of technological and legal indications and the significance of a legal citation decays by time.

To the best of the authors' knowledge, this work represents the first attempt to explore the insight that a patent citation could be a mixture of technological and legal citations, to quantify the technological quality measures and legal quality measures corresponding to the two different dimensions in patent citations, and to apply them in patent valuation. In summary, our work has made the following major contributions:

- This study aims to exploit the technological and legal interpretation of patent citations and apply them to patent valuation.
- Four different patent quality measures, namely, technological richness, technological influence, legal patent scope and legal blocking power, and the interactions among them are proposed to capture the technological and legal information imbedded in patent citations.
- We propose a probabilistic model that considers a patent citation as a probabilistic mixture of technological and legal citations, with the relative weight on the legal dimension decays by time. We develop an algorithm that captures the interdependence among the four proposed patent quality measures to iteratively derive these measures and learn the model parameters, which are useful for analysis of patents in a firm or a field.
- Using patent renewals as an indicator of patent value, our experiments show that considering both the technological and legal dimensions of patent citations and applying these four patent measures can significantly improve patent evaluation, compared to the current practice that only involves the technological interpretation in patent evaluation.

The rest of this paper is organized as follows. We first assume a deterministic model of patent citations and introduce our algorithms to derive measures related to technological citations and legal citations in Section 2. Then we develop a probabilistic mixture model of patent citations to better capture those measures in Section 3. In Section 4, we introduce our evaluation methodology and conduct experiments on valuation of Drug&Medical patents in both firm-level and field-level. We finally review related works in Section 5 and draw conclusions in Section 6.

2. DETERMINISTIC MODEL AND ALGORITHMS

With the technological and legal interpretations for patent citations, an immediate question is how to model and quantify technological and legal citations. In this section, we first discuss a heuristic and deterministic model for interpreting the technological and legal dimensions of patent citations. Then we propose algorithms to quantify the four proposed patent quality measures: technological influence, technological richness, blocking power, and legal patent scope, for a focal patent.

2.1 Modeling of Patent Citations

Here we assume that a patent citation always reflects knowledge flow from the cited patent to the citing patent. Therefore, the technological dimension of a patent citation *always* exists. However, the legal dimension of a patent citation only exists when the cited patent is not expired. In other words, if the cited patent is not expired when the citation is made, the citation is of both a legal and technological citation. Otherwise, the citation only reflects the technological (knowledge) flow. Consider an example where patent A is granted in year 2000 and was maintained (renewed) at its 4^{th} year renewal but not at the 8^{th} year. If a patent B cites patent A in year 2006, we consider the citation to have both legal and technological interpretations. However, for a patent C citing patent A in year 2009, the citation is only a technological citation. Accordingly, it is fairly easy to determine whether a patent citation is a legal citation or not, since we know when a patent is expired (based on the data on patent maintenance at *USPTO*).

Citation Graphs. Based on the discussion above, we derive two citation graphs. One represents the technological citation network, denoted as $G_T = (V, E_T)$ where G_T is the same as the original patent citation network because here we assume that a patent citation always serves its technological functionality, i.e., $E_T(i,j)$=1, if p_i is cited by p_j.[1] On the other hand, the legal citation network, which captures the legal implication between patents, is denoted as $G_L = (V, E_L)$ where $E_L(i,j) = 1$ if p_i is not expired at the grant year of p_j; and $E_L(i,j)$=0 if p_j is expired when p_j is granted.[2] Based on these two citation graphs, we propose to characterize a patent with four quality measures: technological influence score, technological richness score, blocking power score and legal patent scope score. In the following, we describe two basic approaches in quantifying the four features: one is the *CiteCount* algorithm and the other is the *CiteNet* algorithm.

2.2 CiteCount Algorithm

The CiteCount algorithm, similar to the conventional citation counting approach for assessing the quality of scientific literature, counts the number of citations of different types, based on the technological and legal citation graphs. Given the technological citation graph G_T and the legal citation graph G_L, we formally define the Technological Influence score (**TI**), Technological Richness score (**TR**), Blocking Power score (**BP**) and Legal Patent Scope score (**LS**) of a patent p_i as follow.

$$TI_i = \sum_j E_T(i,j) \tag{1}$$

$$TR_i = \sum_j E_T(j,i) \tag{2}$$

where $E_T(i,j)$ refers to a technological citation to patent i made by patent j. Therefore, for patent p_i, its technological influence score is the number of forward citations it receives, while its technological richness score is the number of backward citations it makes.

$$BP_i = \sum_j E_L(i,j) \tag{3}$$

$$LS_i = \gamma - \sum_j E_L(j,i) \tag{4}$$

where $E_L(i,j)$ refers to a legal citation to patent i made by patent j and γ is the global value of legal scope in all patents. Thus, for patent p_i, its blocking power score is the number of legal citations to p_i when it is not expired, while its legal scope score is dependent on the number of citations which it makes when the cited patents are not expired. With p_i making more legal citations, its legal scope is likely further narrowed. Therefore, the resulting legal scope is reduced from the original legal scope of p_i by its legal citations. Two issues arising here are (i) the setting of the initial (global) legal scope value and (ii) the amount of the legal scope to be deducted from this patent due to legal citations. We consider legal patent scope of patent i, as defined by Eq. (4), to be greater than 0. Thus, γ is inherently greater than the maximal number of legal citations/references made by patent p_i. While in this paper all patents are assumed to have the same initial legal patent scope, we may empirically set appropriate γ and use it to analyze the initial legal scopes in different domains or patent sets.

2.3 CiteNet Algorithm

As CiteCount only counts on one-hop neighbors in the graph, the potential influence of neighbor patents located in multi-hop neighborhood is not considered. As a result, the relationship and interdependence between technological influence and richness as well as between blocking power and legal influence are not well considered. To address this issue, we derive CiteNet, a patent citation network based algorithm, to capture these mutual independence. In the algorithm, we make the following intuitive assumptions:

1. The technological influence of a patent is determined by the number of technological citations it receives and the technological richness of these citing patent. The technological influence of a patent will be higher if it is cited by subsequent patents of higher technological richness, as it leads to follow-up innovations of high technological quality.
2. The technological richness of a patent depends on the number of technological citations it makes and the technological influence of these cited patents. The technological richness of a patent will be higher if it cites patents of higher technological influence because it is based on prior innovations of high technological impacts.
3. The blocking power of a patent depends on the number of legal citations it receives and the legal patent scope of these citing patents. The blocking power of patent will be greater, if it is cited by patents with larger legal patent scope because it blocks patents with broader scope.
4. The legal patent scope of a patent is determined by the number of the legal citations it makes and the blocking power of these cited patents in the legal citations.

[1] V is the set of patents and E_T is the set of edges in G_T.
[2] E_L is the set of edges in G_L.

Intuitively, the legal patent scope of a patent will be smaller if it cites a lot of prior patents with stronger blocking power, because these cited patents would narrow down its scope.

Therefore, given a citation network graph $G = (V, E)$, the CiteNet algorithm computes the four quality measures for each patent iteratively until the derived measures converge. In each iteration, the measures are derived as follows.

$$TI_i \leftarrow \sum_j TR_j \quad \text{where } E_T(i,j) = 1 \tag{5}$$

$$TR_i \leftarrow \sum_j TI_j \quad \text{where } E_T(j,i) = 1 \tag{6}$$

Note that Eq.(5) corresponds to the first assumption and Eq.(6) corresponds to the second assumption. Moreover, given the legal citation network G_L and a patent p_i, we have:

$$BP_i \leftarrow \sum_j LS_j \quad \text{where } E_L(i,j) = 1 \tag{7}$$

$$LS_i \leftarrow \gamma - \sum_j BP_j \quad \text{where } E_L(j,i) = 1 \tag{8}$$

Eq. (7) corresponds to the third assumption and Eq. (8) corresponds to the fourth assumption. Also, similar to the CiteNet method, γ is the initial value of legal scope and we subtract the blocking power of legal citations that p_i cites from its original patent legal scope. Moreover, after each iteration, we normalize the calculated scores for the four quality measures using *2-norm* normalization to guarantee the convergence of the algorithm.

In summary, CiteNet captures the independence between technological influence and technological richness, and blocking power and legal scope in each iteration as shown in Eqs. (5)-(8). The proposed CiteNet algorithm derives the patent technological influence and richness in a way similar to the HITS algorithm that captures the mutually reinforcement between authoritative and hub web pages [13]. On the other hand, the derivation of the patent legal blocking power and patent scope is totally different. As Eqs. (7) and (8) show, they are based on different rules.

3. PROBABILISTIC MODELING

In the previous section, we assume that a citation is always a technological citation. Moreover, depending on whether the cited patent is expired at the time of being cited, the citation could have legal implication on the citing patent. Note that some potential issues may arise with such deterministic heuristics. For example, some patent citations are counted twice (as a technological citation and as a legal citation), whereas others are counted only once (only as a technological citation). This seems to be ad hoc. Is there a better way to model the two dimensions of a patent citation coherently? In particular, as explained earlier, a patent citation is likely to be a mixture of a technological citation and a legal citation, with different weights.

Additionally, the legal measures of a cited patent, corresponding to a patent citation, are assumed to remain constant, whether it is cited by a citing patent granted just a few years later or by another patent granted many years later. We argue that intuitively it may be more reasonable to assume that the legal power of a cited patent varies corresponding to different citing patents granted at different years, given that they are all granted before the expiration of the cited patent. In other words, it is more realistic to assume that the legal implication of a patent citation decays as a function of the lagging years between the cited patent and the citing patent until the cited patent is expired.[3]

Therefore, we propose to adopt a probabilistic approach to model patent citations. We assume that the technological and legal interpretation of a patent citation takes some weights, i.e., their total weight equals one. Meanwhile, we assume that the weight for the legal dimension decays as a function of the lagging years between the cited and the citing patents. Moreover, it becomes zero when the cited patent expires.

With such a probabilistic/mixture model of patent citations, we propose, similar to Section 3, *probabilistic citation count (ProbCiteCount)* and *probabilistic citation network (ProbCiteNet)* to quantify the four quality measures of a patent. ProbCiteNet takes into consideration the citation network structure and the interdependence between the two technological measures and between the two legal measures, while ProbCiteCount does not.

Once the decay behavior of legal citations, the constraint over total weight of legal and technological citations, and the relationships among legal and technological quality measures are properly modeled, we shall be able to model the correlations between the quality measures and the patent value, e.g., by formulating it as a classification problem. Putting all these together allows us to not only to derive quality measures but also learn the model parameters (such as the time dacay parameter) and analyze the importance of the patent quality measures in the classifiers for patent evaluation. Furthermore, learning the various model parameters enable us to study insights about valuation of patent citations in a field or a firm.

In the following, we first detail our approach to model the decay function for legal citations (see Section 3.1) as well as ProbCiteCount (see Section 3.2) and ProbCiteNet (see Section 3.3). Then, we introduce our prediction model for patent value and the learning process for deriving model parameters (see Section 3.4).

3.1 Temporal Decay of Legal Citation Weights

Here we discuss the selection of a decay function to model the temporal decay of legal power. Two candidate functions are exponential decay or linear decay functions. An exponential decay function assumes that the rate of decay is proportional to its current value, while a linear decay function assumes that the rate of decay is constant over time, which is less applicable to our case. Consider a case where patent A (granted in 2000 and to expire in 2016) is cited by patent B in 2002 and by patent C in 2003. Since these inventions are very close in time, they are likely close substitutes to each other in the market. Thus, the one year difference between the citing patents B and C could mean significantly different market values. Consequently, the weights on the legal dimension of the two patent citations could be quite different. However, suppose that patent A is cited by patent D in 2014 and by patent E in 2015, which are very far away from patent A (which is about to expire). In this case, the weights for the legal dimension of the citations correspond-

[3]Here the lagging years refers to the number of years the grant date of the citing patent is lagging behind the grant date of the cited patent.

ing to A to D and A to E should be no much different, because inventions in patent A is fading out of the market when it is cited by patent D and E.

Moreover, the legal weight of a patent citation depends on the power of the cited patent in narrowing down and/or blocking the citing patent. Therefore, the rate at which the weight on a legal citation decays should be correlated with the turnover or product life cycle for the technology field of the cited patent. Hence, in this model, we assume that the technology domain of the cited patent, which reflected by the U.S. patent class of the cited patent, determines the temporal decay pattern, i.e., patents in the same U.S. class share the same temporal decay pattern.

Formally, let the parameter of the decay function for a given U.S. class u be λ_u. Given two patents p_i and p_j where p_i is cited by p_j, we define the weight for the legal dimension of this citation as follows.

$$P_L(p_i, p_j) = \begin{cases} \lambda_u e^{-\lambda_u |t_j - t_i|} & p_i \textit{ is not expired when cited} \\ & \textit{by } p_j \\ 0 & p_i \textit{ is expired when cited by } p_j \end{cases} \quad (9)$$

where t_i and t_j are the grant dates of p_i and p_j respectively, and u denotes the U.S. Class of the cited patent p_i.

As we consider any citation to be a mixture of the technological and legal dimensions (with their total weight equals 1), the technological weight for patent p_j citing patent p_i is defined as follows.

$$P_T(p_i, p_j) = 1 - P_L(p_i, p_j) = 1 - \lambda_u e^{-\lambda_u |t_j - t_i|} \quad (10)$$

As such, Eq. (9) and Eq. (10) govern the weighted technological citation network and legal citation network, respectively. Next, we discuss the modeling of interdependency among the four quality measures with ProbCiteCount and ProbCiteNet.

3.2 Probabilistic Citation Count Algorithm

To present the probabilistic citation count (ProbCiteCount) algorithm, we first introduce an adjacency matrix defined for deriving the four quality measures, based on the weighted technological and legal citation networks.

Adjacency Matrix for Technological and Legal Citations. We use an adjacency matrix A_T to denote technological citations and A_L to denote the legal ones. $A_T(i, j) = P_T(p_i, p_j)$ if p_i is cited by p_j, otherwise $A_T(i, j) = 0$. On the other hand, $A_L(i, j) = P_L(p_i, p_j)$ if p_i is cited by p_j, otherwise $A_L(i, j) = 0$.

Based on A_T and A_L, we define technological influence (**TI**) and technological richness (**TR**) for a given patent p_i as follows.

$$TI_i \leftarrow \sum_j A_T(i, j) \quad (11)$$

$$TR_i \leftarrow \sum_j A_T^T(i, j) \quad (12)$$

where j is bounded to the set of patents in the corpus citing p_i and A_T^T is the transpose matrix of A_T. Next, we define blocking power (**BP**) and legal patent scope (**LS**) for a given patent p_i, based on citation count as follows.

$$BP_i \leftarrow \sum_j A_L(i, j) \quad (13)$$

$$LS_i \leftarrow \gamma - \sum_j A_L^T(i, j) \quad (14)$$

where A_L^T is the transpose of A_L and γ is the global legal patent scope for each patent, defined in Section 2.

3.3 Probabilistic Citation Network Algorithm

Given the weighted technological citation matrix A_T and legal citation matrix A_L defined above, the ProCiteNet algorithm iteratively derives the four quality measures, based on the mutual interdependence among them, as discussed in Section 2. Formally, we define technological influence (*TI*) and technological richness (*TR*) for the given patent p_i as follows.

$$TI_i \leftarrow \sum_j A_T \cdot TR_j \quad (15)$$

$$TR_i \leftarrow \sum_j A_T^T \cdot TI_j \quad (16)$$

In Eq. (15) and Eq. (16), j refers to the patents citing patent p_i and the patents cited by patent p_i, respectively.

Similarly, we define blocking power score (*BP*) and legal patent scope score (*LS*) for the given patent p_i as follows.

$$BP_i \leftarrow \sum_j A_L \cdot LS_j \quad (17)$$

$$LS_i \leftarrow (\gamma - \sum_j A_L^T \cdot BP_j) \quad (18)$$

where γ is the initial value of legal scope for each patent, and j refers to the patents citing patent p_i and the patents cited by patent p_i, respectively.

ProbCiteNet, similar to HITS algorithm [13], considers mutual reinforcement between the technological influence and richness, as well as the blocking power and legal patent scope. However, it is different from HITS in that it operates on weighted technological and legal citation graphs that take into account the two dimensions of patent citations and exponential decaying in weights. The weights on the citation networks are to be learned in an integrated learning process, which uses different rules for updating different features based on mutual interdependence among features.

3.4 Prediction Model for Patent Valuation

We argue that the four legal and technological measures can be used for patent valuation and thus derive a model to predict patent value using the proposed quality measures as features. In this section, we introduce a prediction model, which we combine with the probabilistic modeling of patent citations to learn model parameters and to derive classifiers for patent evaluation in an integrated learning process. We use logistic regression model for predicting patent value, and maximize the object function with gradient ascent method to predict patent value. Accordingly, each patent p_i in the training sample can be represented as (x, y) where x is the feature set of p_i. i.e., the four features/quality measures, and y is the predictive label used in regression. For example, the patent maintenance status may serve as a label/indicator of patent value (as discussed in detail later in Section 4). For logistic regression, we define our hypothesis function $h_\theta(x)$ as follows.

$$h_{\omega,\lambda}(x) = \frac{1}{1 + exp(-\omega^T x(\lambda))} \quad (19)$$

where the model parameter θ consists of (i) ω – the weights of the features and (ii) $\lambda = \{\lambda_1, \lambda_2, ..., \lambda_n\}$ – the parameters of the exponential decay functions corresponding to US patent classes. Thus, each feature value in the feature set x is a function of λ.

The objective function to learn the parameter of our model then is set as maximizing the following likelihood function:

$$l(\omega, \lambda) = \sum y^i log h_{\omega,\lambda}(x) + (1 - y^i) log(1 - h_{\omega,\lambda}(x)) \quad (20)$$

where i is the index of the samples in the training data set and y^i is the predictive label for the i^{th} sample.

Learning in ProbCiteCount. By computing the first-order partial derivatives for ω and λ based on Eq. (20), we obtain:

$$\frac{\partial}{\partial \omega_i} l(\omega) = (y - h_\omega(x))x_i \quad (21)$$

and

$$\frac{\partial}{\partial \lambda_i} l(\lambda) = (y - h_\lambda(x))Z_\lambda \quad (22)$$

where Z_λ is the derivative of $\omega^T x(\lambda)$ with respect to λ corresponding to technological influence, technological richness, blocking power and legal patent scope. Basically Z_λ is shown as below.

$$\begin{aligned} &\omega_{ti} \sum_{A_T(i,j)\neq 0} (1 - \lambda_i |t_j - t_i|) e^{-\lambda_i |t_j - t_i|} \\ &+ \omega_{tr} \sum_{A_T(j,i)\neq 0} (1 - \lambda_j |t_j - t_i|) e^{-\lambda_j |t_j - t_i|} \\ &+ \omega_{bp} \sum_{A_L(i,j)\neq 0} (\lambda_i |t_j - t_i| - 1) e^{-\lambda_i |t_j - t_i|} \\ &- \omega_{ls} \sum_{A_L(j,i)\neq 0} (\lambda_j |t_j - t_i| - 1) e^{-\lambda_j |t_j - t_i|} \end{aligned} \quad (23)$$

where ω_{ti}, ω_{tr}, ω_{bp} and ω_{ls} are weights for technological influence, technological richness, blocking power and legal patent scope, respectively.

Therefore, it gives us update rules as follows.

$$\omega_j = \omega_j + \alpha(y - h_\omega x(\lambda))x_j \quad (24)$$

$$\lambda_j = \lambda_j + \alpha(y - h_\lambda x(\lambda))Z(\lambda) \quad (25)$$

where α is the ascent step size.

With the partial derivative with respect to ω, λ solved, we can update ω, λ to maximize the objective function.

Learning in ProbCiteNet. The update process of ω in ProbCiteNet is similar to that in ProbCiteCount. Due to space limitation, we skip this part and only discuss the update process of λ. In order to maximize the objective function, we compute the first-order partial derivatives for ω and λ based on Eq. (20) to obtain:

$$\frac{\partial}{\partial \omega_i} l(\omega) = (y - h_\omega(x))x_j \quad (26)$$

and

$$\frac{\partial}{\partial \lambda_i} l(\lambda) = (y - h_\lambda(x))Z_\lambda \quad (27)$$

where Z_λ is the derivative of the measures with respect to λ. This model calculates four scores iteratively. Suppose that we perform n iterations of updates for the scores, Z_λ is:

$$\begin{aligned} &\omega_{ti}[(A_T^T A_T)^{N-1} + \sum_{n\in N} (A_T^T A_T)^n](\frac{\partial A_T^T}{\partial \lambda_j} A_T + \frac{\partial A_T}{\partial \lambda_j} A_T^T) \\ &+ \omega_{tr}[(A_T A_T^T)^{N-2} + \sum_{n\in N-1} (A_T A_T^T)^n](\frac{\partial A_T}{\partial \lambda_j} A_T^T + \frac{\partial A_T^T}{\partial \lambda_j} A_T) \\ &+ \omega_{bp}[(A_L^T A_L)^{N-1} + \sum_{n\in N} (A_L^T A_L)^n](\frac{\partial A_L^T}{\partial \lambda_j A_L} + \frac{\partial A_L}{\partial \lambda_j} A_L^T) \\ &- \omega_{ls}[(A_L A_L^T)^{N-2} + \sum_{n\in N-1} (A_L A_L^T)^n](\frac{\partial A_L}{\partial \lambda_j} A_L^T + \frac{\partial A_L^T}{\partial \lambda_j} A_L) \end{aligned} \quad (28)$$

where ω_{ti}, ω_{tr}, ω_{bp} and ω_{ls} are weights for technological influence, technological richness, blocking power and patent legal scope, respectively.

4. EVALUATION USING PATENT VALUATION

We conduct experiments to validate our proposition that patent citations have legal and technological dimensions in interpretation and such distinctions can be useful for patent valuation. Specifically, we test whether the proposed patent quality measures (i.e., technological influence and richness, and blocking power and legal patent scope) can improve performance in predicting patent value. In this section, we first discuss our evaluation methodology (i.e., using patent renewal status as a proxy for patent valuation), then conduct experiments to evaluate the effectiveness of the proposed quality measures on patent valuation at individual firms, and finally evaluate the overall performance of patents in Drug&Medical as a whole.

4.1 Evaluation Methodology

To conduct an experimental evaluation on our proposition and proposal using patent valuation, we first need to have some ground truth for patent value. However, it is very difficult to obtain the monetary value of a large sample of patents. Thus, in our experiments, we use patent maintenance/renewal status as an indicator of patent value. After December 12, 1980, US patent holders have to pay maintenance fees for granted patents. Thus, the patent holders need to decide whether to pay the maintenance fees to renew and maintain their patent rights at the 4^{th}, 8^{th} and 12^{th} years after patent grant. Notice that the maintenance fees increase for later renewals. Thus, from the perspective of patent holders, it is intuitive that patent renewal status reflects patent values. For example, a patent that has only a 4^{th} year renewal is less valuable than another patent that is renewed at both the 4^{th} and the 8^{th} years. We assume that patent holders have significant insights about the true value of their patents and therefore use patent maintenance status as a proxy for patent value.

Formally, we define the time period from a patent's grant date to its abandon/expiration date as its *ActivePeriod (AP)*. Given a patent p, AP_p is either 4 years, 8 years, 12 years or more, based on how many times it gets renewed by the patent owner. Therefore, for a given patent p, we use three indicators for its value based on AP of p as follows.

$$C_1(p) = \begin{cases} 1 & if\ AP_p > 4 \\ 0 & otherwise \end{cases} \quad (29)$$

$$C_2(p) = \begin{cases} 1 & if\ AP_p > 8 \\ 0 & otherwise \end{cases} \quad (30)$$

$$C_3(p) = \begin{cases} 1 & if\ AP_p > 12 \\ 0 & otherwise \end{cases} \quad (31)$$

Feature Sets. We use some meta information in patents, including number of patent claims, number of patent independent claims, number of patent inventors, patent U.S. class number, number of patent references, and number of patent citations, as the baseline features for evaluation of patent value. To validate our idea, we focus on the proposed four patent measures (quantified using our CiteCount, CiteNet, ProbCiteNet, and ProbCiteCount algorithms) as features for comparison. For a given patent, we compute the four proposed patent features at the year of patent grant, three years after patent grant, and six years after patent grant, respectively. Right after patent grant, the patent has yet received any citations so we only have information about its backward citations (the prior patents it cites). At the 3^{rd} year after patent grant, the patent has received some forward citations and thus the citation network is quite different. At the 6^{th} year after patent grant, we have more information about its forward citations and the citation network becomes even richer. Again, for different citation networks, we use patent quality measures derived by various algorithms to test whether our findings are robust across various citation networks at different points in time.

Data Set and Experiment Setup. We conduct experiments using patents in the field of the Medical&Drug, collected from the USPTO database, which contains rich text and bibliographic information of US patents. Complete patent maintenance information is also collected. We focus on patents granted between 1995 to 2000, since it takes 12 years to know if a patent has been renewed in all the three renewal decisions. The patent citation networks based on which we quantify the proposed patent quality measures involve more than 100,000 US patents.

We build up three binary classifiers for the three aforementioned binary indicators of patent value, C_1, C_2 and C_3, respectively. These classifiers identify patents of very low value (with an active period of smaller than 4 years) from the rest, patents of above-average value from those of below-average (whether the active period is greater or smaller than 8 years), and patents with high value (with an active period of a full patent life) from others, respectively. We use *Support Vector Machine (SVM)*, a widely used supervised learning model, to build up prediction models and perform a 5-fold cross validation on the data. The prediction accuracy is based on the average accuracy of the binary classifiers for training data in the 5-fold cross validation. Moreover, in order to demonstrate the effectiveness of our proposed features, we use oversampling techniques to construct a balanced data set where the numbers of maintained patents and abandoned patents are the same for each classifier.

4.2 Firm-Level Patent Valuation

In this section, we first evaluate the proposed quality measures and algorithms in predicting patent value for individual firms. We select the top five patent holders in the field of Drug&Medical in terms of the number of their patents issued. This selection covers different types of firms, including pharmaceutical (Merck and Eli Lilly), chemicals (Poctor & Gamble), medical devices (Medtronic), and academic institutions (Regents of the Univ. of California). We analyze model parameters learned for each firm to gain insights in the importance of the four proposed quality measures in deciding patent value in firms. Moreover, we demonstrate and analyze the temporal patterns of decays of legal citations in Drug&Medical.

Data Description. Table 1 shows the maintenance information of the selected firms, where $PSize$ is the number of patents held by each firm, while *Exp.*@4^{th}, *Exp.*@8^{th}, and *Exp.*@12^{th} represent the ratio of patents expiring at the 4^{th}, 8^{th} and 12^{th} year, and *Always Maintained* represents the ratio of patents remaining maintained at the 12^{th} year.

4.2.1 Prediction Results

Figure 1 shows the prediction results on patent value using different feature sets for the selected firms. As shown, we test different combinations of features, with the four proposed patent features derived using different algorithms, based on (forward) citation information at different time points. The results show that incorporating our proposed features (by distinguishing legal and technological citations) into classifier does achieve significantly higher accuracy in predicting patent value than using only the baseline feature set. Recall that these five patent holders vary significantly in terms of their strategy in patent renewal, but our results are very robust across them.

Among the feature sets obtained using different algorithms, we observe that CiteNet achieves a better performance than CiteCount and ProbCiteNet outperforms ProbCiteCount. These results confirm our argument that mutual interdependence and reinforcement between technological influence and richness as well as between legal blocking power and patent scope, which are ignored by the CiteCount and ProbCiteCount. Moreover, comparing the results of *CiteCount* versus *ProbCiteCount* and those of *CiteNet* versus *ProbCiteNet*, respectively, we can see that our proposed probabilistic model achieves a better performance. These results validate the mixture of technological and legal components in patent citations, as we postulated, and the proposed temporal decay in legal citations.

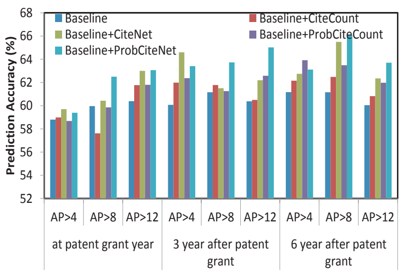
The performance of ProbCiteNet is the best in predicting patent value. This suggests that by capturing both the probabilistic mixture of legal and technological citations in citation networks and the mutual interdependence among patent features, we can achieve the best performance. From the results, we can observe that given no forward citation information, ProbCiteNet improves on the baseline model by 5.95% on average; with three year forward information, the improvement is 9.73%; and given six year forward citation information, an improvement of 10.71% on average.

4.2.2 Analysis of Feature Importance

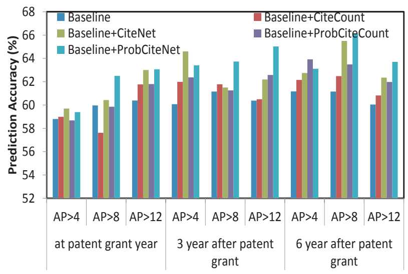
Table 2 shows the learned importantce for different patent quality measures in the prediction model. These measures are derived from the ProCiteNet model based on the citation networks at different time points, i.e., 0-year, 3-year and 6-year after patents are granted. All the importance for these measures is normalized to be between 0 and 1. Note that the importance of a measure reflects the extent of attention received from the patent holder in patent valuation. If the learned predictive model of a company has the largest importance for blocking power, it may indicate

Company Name	PSize	*Exp.*@4^{th}	*Exp.*@8^{th}	*Exp*@12^{th}	*Always Maintained*
Merck	1246	0.427	0.239	0.169	0.165
Poctor&Gamble	1151	0.175	0.150	0.114	0.561
Medtronic	1043	0.038	0.053	0.066	0.842
Eli Lilly and Company	1039	0.119	0.409	0.263	0.207
University of California	963	0.036	0.154	0.278	0.531

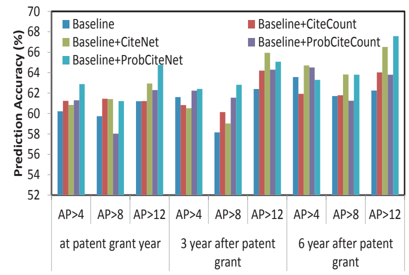
Table 1: Ratio of different maintenance decisions for each firm.



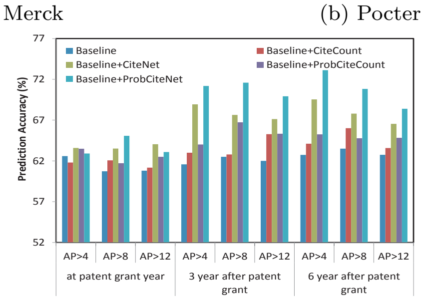
(a) Merck



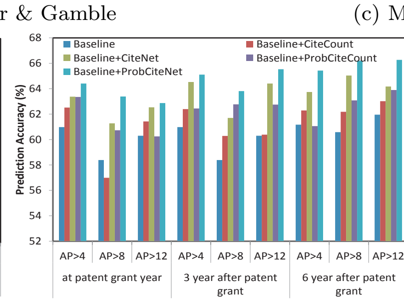
(b) Pocter & Gamble



(c) Medtronic



(d) Eli and Lily



(e) University of California

Figure 1: Firm-Level Patent Valuation. In the figure, "at the patent grant year" means that the feature set is obtained based on citation networks built up at the patent grant year. Similarly, "3 year after patent grant" and "6 year after patent grant" mean that feature sets are quantified based on citation networks built up after 3 years and 6 years after patent grant year.

that the company tends to maintain patents that may block lots of subsequent patents. From the table, we observe that legal quality measures (instead of technological quality measures) play a dominate role in deciding patent value for all the firms under examination, except for the time point when no forward citation is made.[4] In patent valuation, these top firms seem to be focused on the legal implication of their patents, which is an important reason for filing patents.

4.2.3 Analysis of Temporal Decay Parameter

The λ parameter in the exponential decay function, used to model the temporal decay of legal citations in our probabilistic model, represents the speed of technological update in different fields and thus can provide insightful information about these fields. For example, if the technological change in a field is very fast, the techniques and products could be replaced quickly. As a result, a strategy for long-term legal blocking of subsequent inventions in such a field may not be effective as the market/competitors may lose interests in the techniques and products quickly.

To provide more insights, we look into the λ parameter to see if they reflect the speed of technological updates in fields. We select several major U.S. classes, including 424 (Drug, Bio-affecting and body treating compositions), 128 (Surgery), 600 (Surgery)[5] and 435 (Chemistry: Molecular biology and microbiology) in Drug&Medical for this analysis. Figure 2 plots the decay parameters for the aforementioned classes of patents in each firm. The bars for a firm are the average decay parameters corresponding to the maintenance decision at the patent grant date, 3-year after the patent grant, and 6-year after the patent grant, respectively. In each bar, the λ parameters of examined classes are sorted in descending order (from bottom to top). For example, in the first bar, the decay speed of U.S. Class 600 is the greatest and that of U.S. class 435 is the slowest. In Figure 2, we observe that Class 128 and Class 600 have relatively large decay speed, and Class 424 and Class 435 have relatively slow decay speed. Compared to surgery instruments (Class 128 and Class 600), drugs and inventions related to chemistry in molecular biology and microbiology have a relatively slow turnover. This result is consistent with the notion that surgery instruments tend to have a faster product turnover and shorter life cycle, whereas drugs often have a longer life cycle due to the lengthy animal and clinical trials and the regulatory process. Moreover, Class 435 has the slowest decay speed. One possible reason is that inventions in this class might be related to some fundamentals in biology that could lead to drug discovery, thus having very slow speed of technological update.

[4]This is because the blocking power and technological influence is 0, when there is no forward citation (thus it is meaningless to involve these two measures in comparisons under this scenario).

[5]Class 600 is an integral part of Class 128.

Time Point	0 year after grant of patent				3 year after grant of patent				6 year after grant of patent			
Company Name	BP	TI	LS	TR	BP	TI	LS	TR	BP	TI	LS	TR
Merck	0	0	0.8444	0.1566	0.5046	0.0825	0.3762	0.0367	0.46	0.127	0.333	0.08
Poctor&Gamble	0	0	0.73	0.2703	0.372	0.167	0.4	0.0641	0.33	0.18	0.398	0.091
Medtronic	0	0	0.722	0.278	0.53	0.07	0.232	0.168	0.149	0.56	0.19	0.094
Eli Lilly and Company	0	0	0.83	0.167	0.5	0.105	0.317	0.081	0.379	0.151	0.432	0.037
University of California	0	0	0.82	0.18	0.444	0.278	0.194	0.083	0.4	0.312	0.221	0.065

Table 2: Importance of Proposed Patent Quality Measures. BP, TI, LS and TR denote the blocking power, technological influence, legal scope and technological richness, respectively.

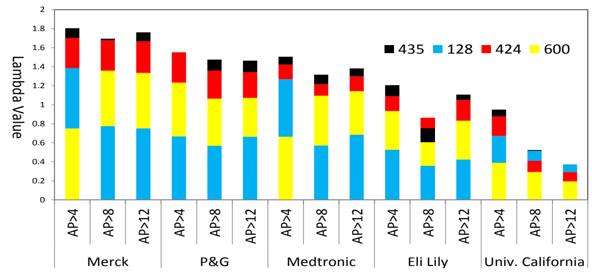


Figure 2: Temporal decay parameter in the probabilistic model: indicating the speed of techniques or ideas being replaced in various fields.

4.3 Field-Level Patent Valuation

After experimenting on firm-level patent valuation, here we aim to analyze how our proposed models perform on the whole set of patents in Drug&Medical. Notice that since patent maintenance fees are paid by each company or organization, we include patent assignee ID in the baseline feature set to reduce the noise caused by varied maintenance strategies by different firms.

4.3.1 Experimental Results and Parameter Analysis

We show the prediction results for the three binary indicators of patent value using the proposed patent quality measures that are obtained from citation networks at various time points (i.e., with no forward citation, 3 years of forward citations and 6 years of forward citations). From Figure 3, we can observe that the overall performance of citation network-based algorithms outperforms those without considering the citation network structure, and using patent quality measures based on probabilistic mixture of technological and legal citations better predicts the patent value. Moreover, the results obtained using different feature sets are consistent with the findings obtained in the firm-level patent valuation experiments (see Figure 1).

However, we can observe that the performance of the overall prediction is not as good as that of patent valuation in individual firms. This is because the patent maintenance decisions are made by individual firms. Although we include the company ID in the baseline feature set, the training data, with patents from different firms, may not fit very well due to the varied maintenance strategies in different firms.

Nevertheless, the analysis on the importance of different patent quality measures still reveals the general pattern in deciding patent valuation in the field. From the learned model, we observe that the blocking power and patent legal scope still play dominating roles in determining patent value (with an average weight of 0.402 for blocking power, 0.327 for legal scope, 0.179 for technological influence, and 0.09 for technological richness).

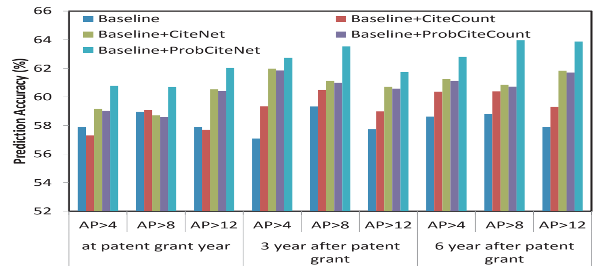


Figure 3: Field-Level Patent Valuation.

Finally, we analyze the decay speed for different classes in the field of Drug&Medical. We observe that the decay parameter λ (which reflects the speed of technology replacement) of U.S. class 600 (Surgery) is 0.67 on average, and U.S. class 435 (Chemistry: molecular biology and microbiology) has the smallest decay parameter (lowest decay speed) of 0.036, which is consistent with the results in firm-level patent evaluation (see Table 2) and our analysis in Section 4.2.3.

5. RELATED WORK

In this section, we review several research lines related to our work, including patent quality assessment, patent novelty detection, patent-relate decision recommendation, and prior art search.

Patent Quality Assessment. Existing works assess patent quality from various aspects, including court ruling decisions [15, 14] and citation counts [19, 1, 4]. There are works in economics which examine the correlations between blocking citations and patent value [4, 1]. Different from previous work, we propose effective algorithms to capture patent quality measures based on distinguishing legal citations from technology citations.

Patent Novelty Detection. Another line of works related to our project is patent novelty detection. Some existing works perform patent novelty detection by analyzing patent claims [9, 3] and calculating the age (the time difference between the first appearance of the word and the current timestamp) of important words in patent claims. Clustering methods are also applied in patent novelty detection. In [17], documents are clustered by their word bags. A document is important if there are lots of work published after it and few works before it in the cluster. [11] ranks a patent based on the activeness of the topics in the patent along the time period.

Patent Decision Recommendation. Different organizations make different patent-related decisions. For a patent examiner, she needs to decide if to grant a patent; for a patent holder, it needs to make the patent maintenance decision. An important line in patent data mining research is to make recommendations for different decision makings. [10] works on the problem of patentability prediction, i.e., to predict if a patent will be granted. Using some extracted contextual features (e.g., word frequency, word age and syntactic complexity), it makes recommendations for patentability decision. [12] works on patent maintenance recommendations, by extracting a set of patent features (e.g., number of patent citations and patent writing quality) to make maintenance recommendations.

Patent Prior Arts Search. Another line of patent-related research is prior art search, which aims to assist inventors and examiners to find relevant patents by exploring similarity between the a given patent application and a set of candidate patents or the citation relationships among patents [5, 18, 2]. Different strategies of enhancing prior art retrievability are studied, including using patent citation [18, 5], PageRank [16] or HITS [13] to improve recall, and formulating the prior arts search problem as a citation recommendation problem [5].

6. CONCLUSION

Patents are important assets for firms in the current knowledge economy, as they protect technologies and markets for patent holders. Therefore, it is crucial for companies to perform effective patent valuation. Patent citations have been widely used in patent valuation. However, unlike paper citations, patent citations not only reflect knowledge spreading but also carry *legal* implication. Based on this observation, we propose and develop algorithms to quantify four patent quality measures that distinguish legal citations from technological citations. In our experiments, we use patent maintenance as an indicator of patent value and predict patent maintenance status effectively. In particular, we propose an integrated process that puts together multiple models (including a temporal decay function on significance of legal citations, a probabilistic citation network based algorithm (ProbCitNet) and a prediction model for patent valuation) and learn the parameters in all these models together in an integrated fashion. Valuable insights, such as decay speed of different U.S. classes, and companies' focus on legal aspects versus technological aspects when it comes to patent evaluation, are found.

This is the first study that differentiates the two dimensions of patent citations and applies this notion to patent valuation. As such, this research points to many promising subsequent research directions. It would be interesting to further study patent valuation in different fields, i.e, drug/biomedical versus information technologies to understand better how legal citations influence patent value. Another challenge is to use various proxies to represent patent value, e.g., to identify the relation between legal citations and whether patents get licensed/sold or litigated.

7. ACKNOWLEDGEMENTS

This work is supported by U.S. National Science Foundation Grants SMA-1063806 and SMA-1360205.

8. REFERENCES

[1] J. Alcácer, M. Gittelman, and B. Sampat. Applicant and examiner citations in u.s. patents: An overview and analysis. *Research Policy*, 38(2):415 – 427, 2009.

[2] S. Bashir and A. Rauber. Improving retrievability of patents in prior-art search. In *ECIR'10*, pages 457–470, 2010.

[3] Y. Chen, W. S. Spangler, J. T. Kreulen, S. Boyer, T. D. Griffin, A. Alba, A. Behal, B. He, L. Kato, A. Lelescu, C. A. Kieliszewski, X. Wu, and L. Zhang. Simple: A strategic information mining platform for licensing and execution. In *ICDM '09*, pages 270–275, 2009.

[4] D. Czarnitzki, K. Hussinger, and B. Leten. The market value of blocking patent citations. ZEW - Centre for European Economic Research Discussion Paper 11-021, 2011.

[5] Y. Guo and C. P. Gomes. Ranking structured documents: A large margin based approach for patent prior art search. In *IJCAI '09*, pages 1058–1064, 2009.

[6] B. Hall, A. Jaffe, and M. Trajtenberg. The NBER patent citation data file: Lessons, insights and methodological tools. *NBER working paper 8498*, 2001.

[7] B. Hall, A. Jaffe, and M. Trajtenberg. Market value and patent citations. *The RAND Journal of Economics*, 36:16–38, 2005.

[8] D. Harhoff, F. Narin, F. M. Scherer, and K. Vopel. Citation Frequency And The Value Of Patented Inventions. *The Review of Economics and Statistics*, 81(3):511–515, 1999.

[9] M. A. Hasan, W. S. Spangler, T. D. Griffin, and A. Alba. Coa: finding novel patents through text analysis. In *KDD '09*, pages 1175–1184, 2009.

[10] S. Hido, S. Suzuki, R. Nishiyama, T. Imamichi, R. Takahashi, T. Nasukawa, T. Idé, Y. Kanehira, R. Yohda, T. Ueno, A. Tajima, and T. Watanabe. Modeling patent quality: A system for large-scale patentability analysis using text mining. pages 655–666, 2012.

[11] P. Hu, M. Huang, P. Xu, W. Li, A. K. Usadi, and X. Zhu. Finding nuggets in ip portfolios: Core patent mining through textual temporal analysis. In *CIKM'12*, pages 1819–1823, 2012.

[12] X. Jin, S. Spangler, Y. Chen, K. Cai, R. Ma, L. Zhang, X. Wu, and J. Han. Patent maintenance recommendation with patent information network model. In *ICDM '11*, pages 280–289, 2011.

[13] J. M. Kleinberg. Hubs, authorities, and communities. *Acm Computing Surveys*, 31:5, 1999.

[14] Y. Liu, P. yun Hseuh, R. Lawrence, S. Meliksetian, C. Perlich, and A. Veen. Latent graphical models for quantifying and predicting patent quality. In *KDD'11*, pages 1145–1153, 2011.

[15] R. Mann and M. Underweiser. A new look at patent quality: Relating patent prosecution to validity. *American Law and Economics Association Annual Meetings*, 2011.

[16] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, 1999.

[17] B. Shaparenko, R. Caruana, J. Gehrke, and T. Joachims. Identifying temporal patterns and key players in document collections. In *TDM '05*, pages 165–174, 2005.

[18] S. Tiwana and E. Horowitz. Findcite: automatically finding prior art patents. In *PaIR '09*, pages 37–40, 2009.

[19] M. Trajtenberg. A penny for your quotes: Patent citations and the value of innovations. *RAND Journal of Economics*, 21(1):172–187, Spring 1990.